

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s): James M. Sweet et al.

Application No.: 10/608,590

Filed: 6/27/2003

Examiner: Nathan Hillery

Art Unit: 2176

Confirmation No.: 8455

Title: DETERMINATION OF MEMBER PAGES FOR  
A HYPERLINKED DOCUMENT WITH LINK AND  
DOCUMENT ANALYSIS

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

**DECLARATION UNDER 37 CFR §1.132**

I, **Steven J. Harrington, Ph. D.**, do hereby declare and state:

1. I am one of the inventors listed on the above-identified application. I reside at 251 Burnett Road, Webster, New York 14580.
2. I have Bachelors degrees in mathematics and physics from Oregon State University in Corvallis Oregon, received in the year 1968, and Masters degrees in physics and computer science from the University of Washington, in Seattle Washington, received in year 1969 and year 1976, respectively. I then also received a Ph.D. in physics from the University Of Washington, in Seattle Washington in year 1976.
3. I have been employed by Xerox Corporation as a scientist and inventor for 26 years, where my current title is Research Fellow, in the Xerox Research Center Webster of the

Xerox Innovation Group. Since I joined Xerox, I have performed research and development in the area of document engineering and digital imaging technologies and have been granted over 120 patents in those technology areas.

4. I have read and understand the material provided with regard to patent application No. 10/608,587 by Sweet et al. including the application itself, the amended claims, patents 6,112,203 to Bharat, 5,924,104 to Earl and 6,877,002 to Prince, as well as the remarks and arguments to the patent examiner the contents of the U.S.P.T.O. Official Action of August 3, 2007, and am of the position that a) claims 1-6, 10-13, 16-20, 25-31, 36, and 37 are not obvious over Bharat et al. (U.S. Patent 6,112,203) in view of Earl, (U.S. Patent 5,924,104), and that claims 7-9, 14, 15, 21-24 and 32-35, are not obvious over Bharat et al., in view of Earl, in view of Prince, (6,877,002).

5. The teachings of Earl deal with intra-page links even though they are referred to as intra-document links. As such they would be removed from consideration by the method as taught and claimed in the present Application. The Earl patent does nothing to provide enlightenment towards the problem addressed by the present Application. Nor does the Prince patent provide teachings that address the shortfalls of Bharat and Earl. Bharat is attempting to assemble a set of distinct documents relevant to a search topic, while the present Application is teaching the identification of links to components of a single document. As such, the Bharat method quickly discards the very links that the present Application is seeking to identify. Furthermore, since Bharat's method rejects many documents and web pages besides those belonging to a document, one could not simply collect the referenced pages that Bharat rejects. Some additional points are that Bharat is working from a set of documents identified by a search engine and therefore has no specified preferred document for use in the identification of its hyperlinked components. Furthermore, since Bharat is working on a set of document identified by a search engine, there is not guarantee of connectedness of the resulting n-graph. None of the teachings of Bharat do anything to detect or enforce connectedness, while for the problem of document boundary identification addressed by the present Application, connectedness of the graph is central, and the process taught of following links chained from an original

source page, guarantees it. While both Bharat and the present Application examine the same space of hyperlinked web pages, and both seek to classify web pages identified in that space, their goals are quite different, leading to different search, analysis, and classification techniques. The teachings of Bharat cannot be used to solve the problem addressed by the present Application, and do not anticipate the teachings of the present Application.

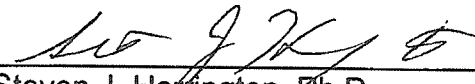
6. With regards to the amendment of the claims as to a document representation "stored in memory". We the Applicants are dealing with exactly the same document representations as are Bharat and Earl, namely web pages provided by a web server. It will be understood by one skilled in the art that a web server is a computing device whose purpose is to deliver web pages to web browser's over a communications channel such as the internet. It will be understood by one skilled in the art that a web server consists of a processor and memory. It will be understood by one skilled in the art that the memory of the web server contains the web page document in a representation from which the processor can copy, transform or otherwise generate the form understood by the web browser that is requesting the web page. The present application claims "an automated identification methodology for assembling document related hyperlinked pages". It further describes and claims "an automated document boundary detection system". As one skilled in the art, it is obvious to me that automation of the method entails a computing system capable of retrieving and analyzing the electronic representations of the web pages in accordance with the methods taught by the application specification. Such a computing system would have a processor able to conduct the analysis and memory to hold the document representation and grouping results. I therefore believe that the amendments to the claims indicating that the resultant set of document pages are grouped into a document representation stored in memory, is supported by the specification in the light of the digital electronic nature of the documents and the indication of automated processing in the claims, and would be so understood as such by those skilled in the art.

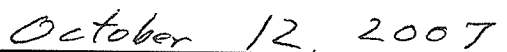
7. As to the term "weed out", this will be understood by one skilled in the art as meaning to discriminate *and discard*. Gardening terms such as "weed out" or "prune" are not uncommon in computer science and are used with usual English understanding of eliminating from further consideration. In computing, this often includes removal of the weeded object from memory, but whether or not this is the case, the weeded object will no longer be included among the objects being processed. This meaning can be seen in the specification which states "...links 240 are passed to module 250 for a final examination to weed out links which have properties that are not characteristic of typical intra-document links..." and then in the next sentence state "The final result is then a list of intra-document links 120 for the candidate page 270". In other words, the links that are not intra-document links are not only distinguished, but also discarded by the weeding process, leaving just the intra-document links.

8. I, the undersigned, further declare that all statements made herein are of my own knowledge are true and that all statements made on information and beliefs are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both under Section 1001 of Title 18 of the United States Code, and further such willful statements may jeopardize the validity of the application or any patent issuing thereon.

Signed:

Respectfully submitted,

  
\_\_\_\_\_  
Steven J. Harrington, Ph.D.

  
\_\_\_\_\_  
Date